DIGITAL dts SOUND

DIGITAL THEATER SYSTEMS
DIGITAL SOUND FOR MOVIES

DOCKET FILE COPY ORIGINAL

RECEIVED
JUL 10 1996
FCC MAIL ROOM

Federal Communications Commission
1919 M Street, N.W.
Washington, D.C. 20554

July 9, 1996

Dear Sirs:

Pursuant to applicable procedures set forth in Sections 1.415 and 1.419 of the Commission's Rules, 47C.F.R. Sections 1.415 and 1.419, Digital Theater Systems wishes to formally file comments at this time.

Enclosed please find one original and eleven copies to be distributed to each of the Commissioners.

Sincerely,

Terry Beard
Chairman, C.E.O.

Enclosures
TB/mac

No. of Copies rec'd 0+11
List ABCDE

31336 VIA COLINAS #101
WESTLAKE VILLAGE, CA 91362 U.S.A
PHONE 818 706 3525
FAX 818 706 1868
PHONE 800 959 4109

Before the
Federal Communications Commission
Washington, DC 20554

In the Matter of	)	
	)	
Advanced Television Systems	)	
and Their Impact Upon the	)	MM Docket No. 87-268
Existing Television Broadcast Service.		

Comments of Digital Theater Systems, LP

Digital Theater System(DTS) hereby submits its comments in response to the Fifth Further Notice of Proposed Rule Making("Fifth Further Notice) adopted on May 9,1996 and released on May 20th,1996 by the Federal Communications Commission("Commission").

In Its Fifth Further Notice the Commission proposed adoption of the ATSC Digital Television Standard. Specifically, the Commission proposes to require the use by digital television licensees of each element of the standard. In these comments DTS recommends that a modification be made to the ATSC audio standard to allow for the inclusion/implementation of alternative audio coding systems, including DTS.

Audio Coding

DTS recommends that the ATSC DTV standard

• **not require conformance with ATSC Doc. A/52, the Digital Audio Compression (AC-3) Standard.**

• **include provision for bitstreams produced by different audio coding systems within the MPEG 2 transport layer, using descriptors/headers.**

• **ensure that all audio coding systems selected for transmission are capable of delivering audio transparently.**

• **ensure that all audio coding systems selected for transmission are capable of operating on a minimum target hardware decoder [1].**

The ATSC DTV audio standard (figure 1) as proposed (AC-3) is already technically obsolete compared to more advanced audio coding systems [2], and is poised to limit further innovation in the broadcast delivery of audio to consumers.

The DTS audio coding system is capable of providing all of the services described the Audio Systems Characteristics of the Digital Television Standard (Annex B). DTS can encode a complete main audio service which includes left, center, right, left surround, right surround, and low frequency enhancement channels into a bitstream at a rate between 320 and 384 kilobits per second (kbps). Furthermore DTS is also capable of operating at the higher bit rates preferred by many content providers. The audio service is not limited to 5.1 channels but can deliver from one

[1] DTV receiver hardware that is capable of decoding the most elementary broadcast bitstream

[2] Technology Update, "Down the stretch , Dolby's still in the lead, but..." Larry Johnson , New York Times 7/7/96

to eight channels over a wide range of bit rates. Multiple audio bit streams may be delivered simultaneously for multiple languages or for services for the visually or hearing impaired. The DTS system also contains features which would allow viewers to control fluctuations in audio level between programs or to select the full dynamic range of the original audio program.

Additionally the DTS coding system offers the advantages listed in Section 2. A more detailed explanation of the DTS coding algorithm is given in, " DTS Coherent Acoustics. Delivering high quality multichannel sound to the consumer", April 1996, presented at the 100th Convention of the Audio Engineering Society, Copenhagen.

Analysis of Required Standards

We believe that adoption of the ATSC Doc. A/52 excludes already existing achievements and will deter further technical innovations in the field of audio coding. As currently written, the Standard locks the digital television broadcast market into a less than optimal, and already obsolete, audio technology.

There is little or no commercial advantage in having a single mandatory audio coding standard for digital television broadcasts. Competing audio coding systems can, and already do, operate on exactly equivalent silicon hardware. In other words, innovations in this area will increasingly resemble software upgrades in the computer market. New and more technically advanced audio decoding algorithms, and other audio processing functions, will be offered as upgrades to the consumer and operate on the existing hardware. This will allow service providers to choose different encoding technologies, and thus compete by offering differentiated products.

To implement this scheme requires that:

- a minimum target hardware decoder, resident in the receiver, be standardized (figure 1)
- unique descriptors and headers be specified within the MPEG-2 transport to facilitate the transmission of alternate audio and video bit-streams
- audio decoding software be downloadable to the hardware decoder via the transmitted data stream.

Response to Proposal

We feel that the audio specifications of the digital television system that have been recommended by the Advisory Committee are somewhat inflexible. Moreover, they do not provide high quality audio.

The proposed audio standard is inflexible in that it does not allow the broadcaster, artist, or producer to select the audio encoding system most appropriate for their material. For example, numerous producers and directors have chosen to encode their film and/or music programming in DTS for theatrical or music distribution. These artists should be given the opportunity to have the same choice when their works appear in the broadcast medium.

The proposed audio standard is also inflexible in that it cannot incorporate rapidly evolving improvements in audio recording techniques including, for example, longer word lengths, higher sampling rates and more audio channels (beyond 5.1 channels). The ATSC standard should include provisions within the audio standard that would allow these improvements to reach the consumer.

Although the AC-3 standard is probably an improvement over today's NTSC monophonic and stereo standards, it is still not comparable to the CD in terms of quality, an audio standard that consumers now expect from all digital media, and which is enjoyed in almost 50% of American homes. [3]

Digital Theater Systems is the premiere motion picture digital sound system, installed in more than 7000 movie theaters worldwide. Beginning with its first release of Universal Pictures Jurassic Park, DTS has positioned itself as the worldwide standard release format for motion picture digital sound tracks. Ten major studios/producers have released over 130 films in the DTS digital sound format. These include Universal, Amblin Entertainment, MGM, Paramount, Miramax, Warner Bro., Castle Rock Pictures, 20th Century Fox, Savoy, and New Line Cinema.

Digital Theater Systems has also developed a real-time consumer audio compression algorithm, Coherent Acoustics. This algorithm was designed from the outset to perform at studio grade quality levels i.e. 'better than CD', and in a multichannel format was intended to facilitate a major advance in the quality of audio reproduction in the home in terms of fidelity and sound stage imagery. Another primary objective was that the compression algorithm should be broadly applicable and therefore flexible. Multimedia applications have restricted data bandwidths and therefore demand a 5.1 channel mode operating at 384 kbps or less. Professional music applications involve higher sampling rates, longer word lengths, multiple discreet audio channels, and increasingly demand lossless compression. All of these features have been accommodated in Coherent Acoustics. The final important objective was to ensure that the universal decoder algorithm was relatively simple, and future-proofed. This would ensure cost effective consumer decoding hardware today, and yet allow consumers to benefit from any future improvements realized at the audio encoding stage.

By not mandating a particular audio encoding scheme within the ATSC DTV standard, broadcasters could take advantage of the flexible nature of the packetized data transport structure of the overall DTV proposal. Within the MPEG 2 transport layer it is quite straightforward to include provisions for many different types of audio coding techniques, including the DTS 'Coherent Acoustics' audio coding system. The advantages of this approach are numerous:

- provides flexibility to artists and broadcasters
- encourages continuing innovation in audio coding techniques
- enables higher quality audio coding
- stimulates competition
- increases product differentiation
- facilitates international compatibility
- enhances international competitiveness of the US standard
- enhances opportunity for US based content providers

Telecommunications Act of 1996

The Proposed ATSC DTV audio standard discourages new entrants into the marketplace. **The Fifth Further Notice of Proposed Rule Making** cites the provision of the Telecommunications Act of 1996 which seeks,

"to promote competition and reduce regulation in order to secure lower prices and higher quality services for American telecommunications consumers and encourage the rapid deployment of new telecommunications technologies. "

By giving a monopoly to a single proprietary audio coding system (AC-3), the proposed audio standard departs from this principle. It discourages competition, creates a climate for higher prices, and fails to ensure that the highest quality audio is provided for consumers. The proposed

[3] The Year in Consumer Electronics 1995, Consumer Electronics Manufacturing Association, " Household Penetration of Consumer Electronic Products as of January 1996", pp.14.

audio standard also stifles "the rapid deployment of new communications technologies" such as that offered by DTS. Our belief is that by mandating a single proprietary audio coding system, the spirit of the Telecommunications Act (1996) will not be realized.

We believe that there should be a choice of audio coding technologies in the standard, and that the digital licensees should be free to choose audio coding and compression systems appropriate for their program material.

Acceptability of the ATSC DTV Standard

We are concerned that the proposed audio standard will place a wall between the merging of Personal Computers and television. DTS offers a clear advantage by allowing software decoders to progressively decode the coded audio bitstream. For example, a schoolroom with older model PC's would be able to listen to the full audio content but with a restricted bandwidth. The intelligibility of the programming would remain intact. This "future proofs " today's hardware and gives consumers with varying economic means access to content.

DTS was conceived as the artists medium and we support a standard that faithfully reproduces the artists intent. Just as arbitrary cropping of a picture is offensive to the director and cinematographer, so a destructive audio codec aggrieves the musician and composer. Quality compromise is not an inevitable feature of coding. Recording engineers and musicians describe "transparency " when there is no audible difference between the master tape and final delivery to the consumer. We believe that transparency is an artists right. DTS opposes imposing a standard that is incapable of reproducing the artist intent, and offers a technology which is capable of more faithfully reproducing the artists intent within the constraints imposed by ANNEX B, Audio Systems Characteristics.

In addition, the validity of the ATSC audio test procedures and the analysis of the results are suspect. We agree with the CRC who stated that,

"the validity of the test results will be questioned by members of the international community since they were not obtained by a test procedure that is fully compliant with test procedures that have been agreed upon internationally." [4]

As the BBC also pointed out the test procedures employed,

" had a profound effect on the criticality of the assessments and particularly the ability of the test listeners to make valid judgments. There are also serious departures from the normal practices of statistical analysis of subjective test results. Both of these factors throw into doubt the validity of the ACTS overall conclusions on the absolute quality of the Grand Alliance Sound System, AC-3". [5]

The ACTS conclusion that the AC-3 coding is transparent for all of the test items is NOT technically supportable. <u>Therefore, we believe the audio standard should incorporate other coding technologies such as DTS that are capable of delivering transparent audio.</u>

[4] Letter from CRC to the chairman of the ACTS SS WP-2 Committee 1/17/95

[5] Comments on the report to the Advisory Committee on Advanced Television Service of the Federal Communications Commission," digital HDTV Grand Alliance System: Record of Test Results; D.J. Meares and D.J. Kirby, BBC Research and Development Department, Kingswood Warren, Tadworth, Surrey, UK.

Interoperability

DTS facilitates cross-industry interoperability because DTS is capable of automatically operating in any mode indicated by the incoming data stream. A second feature that enhances interoperability is the flexibility of the DTS algorithm to operate across all current and future media.

As bandwidth in non-broadcast media expands, the importance of scalability increases. DTS's ability to operate over a wide range of bit rates, and its progressive decoding capacity combine to create a powerful tool for cross-industry operation. These parameters serve to avoid creating a "tower of babble syndrome" and serve to support seamless integration of programming and programming elements. DTS was designed with the idea of eliminating technical incompatibilities in the transfer process in all media downstream and upstream from film/music program material.

Provision for implementing DTS, such as bitstream headers and descriptors and a standardized hardware audio decoder, are additional actions the Commission should take to facilitate interoperability.

We envision a DTV system that will allow a music student to download the real-time broadcast performances of an orchestra, download the score, perform a particular part, and receive "master-class" level instruction. We envision theatrical performances of Shakespeare, the musical performances of Baba Olatunji and newscasts from science laboratories all kindling the imagination of students in vital ways. Good teaching depends on intelligibility, and music on the communication of the subtlest nuances. We believe the encoding tools of DTS allow for the full range of communication as personal computing and broadcasting merge.

Licensing Technology

Patents covering the DTS 'Coherent Acoustics' audio coding system are currently pending. A license will be made available to applicants under reasonable terms and conditions that are free of any unfair discrimination.

International Trade

We believe that our proposal to allow additional audio coding systems within the ATSC DTV standard would

- facilitate international compatibility
- enhance the competitiveness of a US system worldwide
- remove barriers for US produced films
- enhance the opportunities of US based content providers and equipment manufacturers.

It would allow other countries to adopt the ATSC DTV standard, but provide a choice of the audio coding scheme that more appropriately met their needs or national priorities.

Conclusion

It should be noted that the over-all sonic quality of 5.1 channel motion picture sound tracks has risen considerably in the past three years, and will continue to rise. Broadcasters must allow for this general trend, and provide methods by which ongoing improvements in audio production can be delivered to the consumer.

Mandating a single proprietary audio coding system, such as AC-3, excludes already existing achievements, locks the digital television broadcast market into an already obsolete audio technology, and deters further technical innovations in the field of audio coding.

Digital Theater Systems LP, through the use of the Coherent Acoustics audio coding system, can provide excellent sound quality and flexibility today, and offers scalability for future improvements in services as demanded. DTS recommends that the audio standard be made more flexible by facilitating the transmission of a range of alternate audio bit streams, capable of running on a single universal hardware decoder, such that the service provider has the option of tailoring the audio coding system to the program material and/or service.

Section 2
Advantages of DTS compared to AC-3

Higher audio quality
The DTS audio coding algorithm is more efficient than AC-3, resulting in higher quality audio at an equivalent bit-rate; subjective assessments on 1 and 2 channel material demonstrate this [6].

An efficient audio coding algorithm is of particular importance for broadcast applications in three respects:

1. Live broadcasts of 5.1 channel material will demand a higher 'coding margin' than pre-processed 5.1 channel broadcasts. In this context 'coding margin' refers to the noise masking threshold of an audio signal with respect to the coding induced noise level in the signal. For a fully transparent coding scheme the coding margin must always be positive implying that the coding induced noise is always less than the noise mask threshold and hence inaudible. The 5.1 channel AC-3 algorithm is used almost exclusively as a post-production process, where it is possible to monitor the quality of the coded audio and 'redo' the compression cycle if problems arise due, for example, to critical program material. For live broadcasts it is not possible to optimize the coding algorithm in this manner, and the compression algorithm must be able to deliver transparency in a real-time one-pass mode for all program material.

None of the FCC audio sound quality tests used in-line real-time audio coding-decoding hardware, which is of critical importance for live broadcast applications.

2. The use of 5.1 channels for non-film programs will rise. Music programs in particular benefit from the extra surround channels but, in comparison to motion picture sound tracks where the surround channels are used sparingly, musical material places great demands on the continuous high fidelity of every channel. The DTS algorithm operating with 5.1 channels at 384 kbits/sec has been designed to faithfully reproduce all surround programs whether film based or music based.

3. Broadcasters must also be made aware of problems caused by tandem coding whereby audio material in the post production chain is repeatedly data compressed and uncompressed before final transmission to the consumer. These problems are already apparent to radio broadcasters who have been using data compressed audio in production and point-to-point transmissions for at least five years. The potential for audible artifacts caused by tandem coding requires that a 'coding margin' be allowed in the final transmission to the consumer. A high compression coding technique that is working at 'the edge of transparency' may not be sufficient, even in the near future.

---

[6] DTS Coherent Acoustics 'Sound quality evaluation disk'. CD available from DTS Technology LP

More flexible algorithm

Fully independent channels
Due to its higher coding efficiency the DTS algorithm is able to code each channel fully independently at bit rates less than or equal to 384 kbits/sec (for 5.1 channels). This allows multi-lingual broadcasts to be efficiently transmitted by, for example, simply replacing the main dialog channel in motion picture sound tracks. High quality monophonic speech can be transmitted at a bit-rate less than or equal to 64 kbits/sec.

Open ended architecture
The DTS algorithm is extendible in allowing additional audio channels to be added as required. These extra audio channels retain compatibility with the main 5.1 data stream, and may be decoded or ignored depending on the complexity of the decoder. The architecture also permits higher sampling rate audio to be encoded (e.g. 96 kHz) and longer word-lengths to be used (up to 24-bit data samples) while retaining compatibility with basic decoders.

Accurate frame alignment with video signals
DTS incorporates a synchronization routine that enables sample accurate alignment of the compressed video and audio frames. This is of fundamental importance for seamless switching between multiple compressed video streams, and for time aligning multiple compressed audio streams with a common video stream.

Scaleable algorithm
The DTS algorithm is capable of operating at fixed bit-rates ranging from 32 kbits/sec up to 4096 kbits/sec, and also in a variable-rate lossless mode. This range of bit-rates enables a wide range of applications including monophonic voice transmission, 5.1 channel motion picture sound tracks, or original programming of very high quality orchestral music. DTS is suitable for any application.

The bit-stream format also allows a sub-set of the compressed audio data to be decoded as a band-limited audio signal, if the audio decoding processor is not capable of decoding the full data set. In this way the algorithm can be tailored to the cost/complexity of the decoder.

AUDIO SYSTEMS

Audio coding system overview

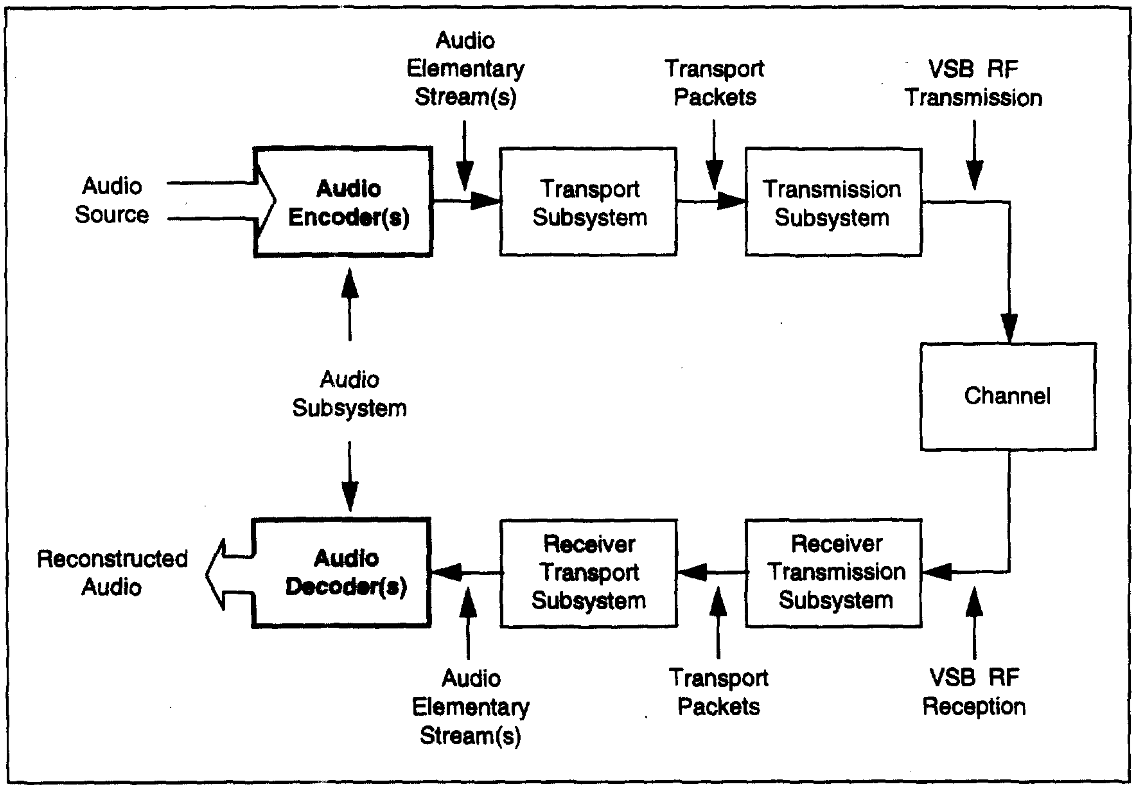


Figure 1. Audio subsystem within the digital television system.

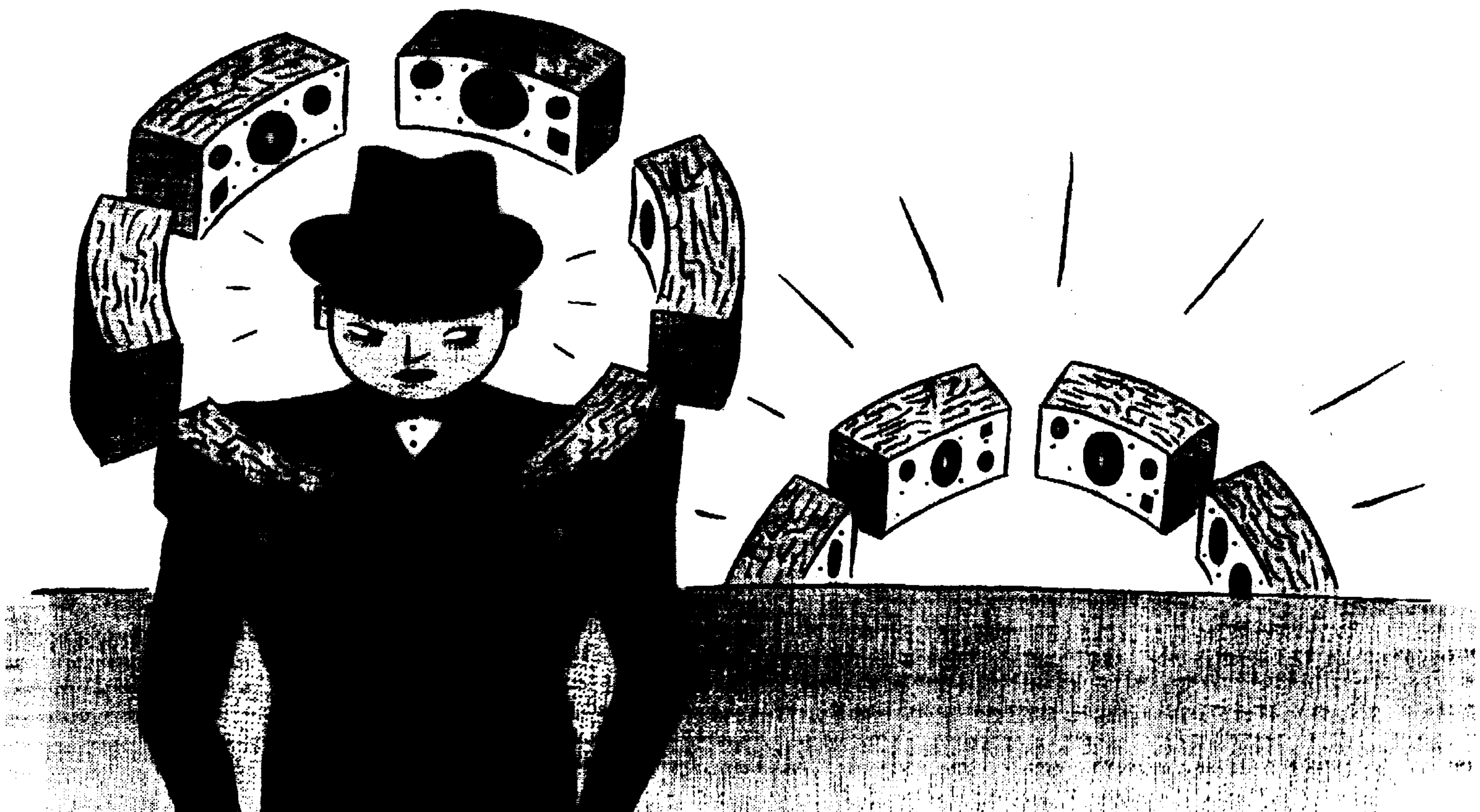
The New York Times

NEW YORK, SUNDAY, JULY 7, 1996

[redacted]

TECHNOLOGY VIEW/Lawrence B. Johnson

Down the Stretch, Dolby's Still in the Lead, but . . .



UNTIL VERY RECENTLY, the nascent technology of digital surround sound had but one name: Dolby. Despite mixed views of its sonic quality, on laser video disks Dolby Digital multichannel recording has been more than dominant; it has been the only game in town. And in setting guidelines for the coming digital video disk, the electronics industry has decreed that the Dolby digital multichannel decoding system must be included in every player.

But now Dolby may find itself in a horse race. The competition, suddenly moving up fast, is Digital Theater Systems, a California-based company whose digital surround scheme is used in more than 3,000 cinemas across the country, nearly three times the number of movie houses equipped with Dolby's commercial digital system.

For all its professional success, Digital Theater Systems was slow to develop a version of its multichannel technology for the consumer market. Even as its consumer system, DTS Coherent Acoustics, underwent refinements, demonstrations were rare. In short, "DTS," as the system is commonly known, persisted as more of a buzz than a reality. It had a few ardent champions but not much of a presence.

That obscurity ended with a bang last month at the Hi-Fi '96 home theater and high-end audio exposition in New York. In parallel demonstrations, DTS not only showcased the vivid, precise surround-sound field its method can bring to movies in a home-listening environment; it also made a powerful case for DTS Coherent Acoustics as a medium for recording music.

At the moment, one can only wait with whetted appetite. While Dolby boasts a number of laser disks bearing its digital multichannel soundtracks (not to be confused with conventional Dolby Surround), the DTS score remains at zero. The first DTS-encoded laser disks — "Jurassic Park," "Apollo 13" and "Casper" probably among them — are promised by September. By Christmas, as many as 20 titles should be available, according to David DelGrosso, the company's marketing director.

Meanwhile, electronics manufacturers are planning their first surround-sound processors with the chip required to decode the DTS signal. Mr. DelGrosso said about a dozen DTS-equipped processors would be available by the end of the year, with the first models expected in August or September.

What proved so impressive in the DTS movie sound at Hi-Fi '96 was not just the whiz-bang effects in the rear channels but also the subtle layering of sound and its untiring character. Listening to goodly stretches of "Jurassic Park," "Apollo 13" and "Casper," movies I've come to know almost by rote, I found myself engaged at a new level of intensity and delight. It was an experience more like cinema than video.

Was it better sound than the best heard from a Dolby Digital laser disk? It was unquestionably more refined, more elegant. One might say more beautiful if that were not so bizarre a word for the roar of rocket engines.

The sense of beauty was only heightened in the second DTS demonstration, which concentrated on music CD's. Not even the most ardent proponents of Dolby Digital have urged that system for music recording. Its high degree of digital compression would take a severe toll on the sound of music. Indeed, the deleterious effect of high-order compression on music has been amply demonstrated by both the digital compact cassette and the mini-disk.

DTS uses a less extreme ratio of compression, and the results seemed plausibly close to what recording engineers like to call transparency: no difference at all between the master tape and the final multichannel CD. A sampling of DTS-encoded CD's (among them the Steve Miller Band's "Fly Like an Eagle" and Bachman-Turner Overdrive's "Not Fragile"), on the High-Definition Surround label, was perhaps the signal event of the entire five-day, far-flung Hi-Fi '96 show.

Buzz becomes reality in a new surround-sound package from Digital Theater Systems, and the race suddenly heats up.

That brief audition afforded a very encouraging glimpse into the future of music recording in the multichannel era. You can count on this: the history of two-channel recording is all but written. It scarcely matters that you can't buy a DTS decoder yet, or that DTS-encoded CD's are incompatible with regular CD players. The needed gear is coming.

The technology itself is the thing, and its far-ranging possibilities will almost certainly affect not only the way music is recorded but also the way it is written. As a full-blown music recording system, DTS will allow the first uncompromised realization of four-channel masterings from the quadraphonic era of the early 70's. The system also presents composers today with a blank slate on which to create multidimensional works, which can be captured in as many as eight channels.

IN A WORD, THE IMPRESSION made by this first rigorous presentation of consumer DTS was stunning. But particularly interesting was the company's retro posture. DTS appears to be committed to the CD in its current form and to the laser disk. And why not, since the major companies behind the digital video disk have given Digital Theater Systems the cold shoulder? Sure, the producers of a particular movie on digital video disk can utilize DTS surround sound, and yes, hardware manufacturers can — if they see a need — include a DTS chip along with the requisite Dolby Digital.

Well, keep an eye on the fast-rising image of DTS, because it has the look and feel of a groundswell. The question may be, who needs whom? In its radical conservatism, DTS is hitching its wagon to twin stars, the CD and the laser disk, that are likely to shine for some time to come. The digital video disk has yet to twinkle.

By the time that medium finally emerges, the electronics industry may have gained a whole new perspective on digital surround sound. The movers and shakers may have shimmied over to DTS. □

DTS Coherent Acoustics. Delivering high quality multichannel sound to the consumer

SMF Smyth, WP Smith, MHC Smyth, M Yan, T Jung
DTS Technology LP, 31336 Via Colinas, Westlake Village, CA 91362

Presented at the 100th Convention 1996 May 11–14 Copenhagen

AES

This preprint has been reproduced from the author's advance manuscript, without editing, corrections or consideration by the Review Board. The AES takes no responsibility for the contents.

Additional preprints may be obtained by sending request and remittance to the Audio Engineering Society, 60 East 42nd St., New York, New York 10165-2520, USA.

AN AUDIO ENGINEERING SOCIETY PREPRINT

be presented at

e 100th AES Convention, Copenhagen 1996
ltichannel Sound Production Techniques and Technologies (Workshop 4a-3)
air Gerhard Steinke

DTS Coherent Acoustics
Delivering high quality multichannel sound to the consumer

SMF Smyth, WP Smith, MHC Smyth, M Yan, T Jung*

DTS Technology LP
31336 Via Colinas, Westlake Village, CA 91362
Tel 818 706 3525 Fax 818 706 0980

*Digital Music Products Inc
Park Square Station, PO Box 15835
Stamford CT 06901

Abstract

h low bit-rate multichannel coders becoming more prominent in the home theatre market, attention is inning to focus on the applicability of the multichannel format for music reproduction in the home. What erences are there between discreet motion-picture sound tracks and multichannel music, and how s this affect the performance of low bit-rate coders ? What are the limitations of the current CD and DVD platforms as multitrack music players, and what impact will these limitations have on the elopment of a new standard format for the release of multichannel music ?

s paper will analyze the potential of DVDs and CDs as high quality multitrack audio media, focusing in icular on the use of DTS Coherent Acoustics as a means of extending the multichannel potential of e platforms. Designed to achieve audio transparency, DTS Coherent Acoustics can operate at bit-s up to 4.096 Mbps, supports up to 8 discreet channels of audio, at up to 24 bit sample resolution and ampling rates up to 96 kHz per channel. Within these constraints DVD can deliver a new multichannel ic experience to the home. On CDs and laser discs DTS has already demonstrated a 6-channel ability, at 20-bit resolution and at a sampling rate of 44.1 kHz. Examples of music recordings made to format by one of the authors will be on demonstration at the workshop.

Introduction
The introduction of discreet multichannel digital playback systems in the home for the reproduction of motion picture soundtracks has rekindled professional and consumer interest in multichannel music. This interest has intensified recently due to the eminent consumer launch of video DVD hardware and software, which offers a much higher bandwidth and greater data capacity than the current CD format.

Limitations of the new DVD platform for multitrack music
At first glance the DVD platform appears to be ideal for multichannel music using linear PCM, but on closer analysis the proposed maximum data throughput of the audio DVD players still restricts the multichannel music capabilities of this new media. This is particularly true if a stereo format is required in addition to the multichannel format, and the stereo format evolves to audiophile quality in terms of sample precision and frequency resolution. A stereo mix may be mandatory since a multichannel to 2-channel mixdown within the DVD player is not considered viable at an artistic level. Table 1 outlines the multichannel possibilities using linear PCM, assuming the inclusion of mandatory 2-channel stereo at 20-bit precision and at a sampling rate of 48 kHz. The maximum audio data rate is assumed to be 6.144 Mbps.

If the stereo format moves to a sampling rate of 96 kHz, then essentially there is no possibility of a multichannel format using linear PCM. Also if low bit-rate video for music videos (e.g. MPEG-1 at around 1.2 Mbps) is included as an option then the audio throughput may drop and further restrict the multichannel linear PCM potential of DVD.

Table 1 Multichannel audio modes assuming mandatory 2-ch PCM @ 20-bit @ 48 kHz (1920 kbps)

sample rate [kHz]	word length [bits]	maximum number of linear PCM audio channels for multichannel format									multichannel bit rate [kbps]	total bit rate [kbps]
48	16	x	x	x	x	x					3840	5760
48	20	x	x	x	x						3840	5760
48	24	x	x	x							3456	5376
96	16	x	x								3072	4992
96	20	x	x								3840	5760
96	24	x									2304	4224

Digital audio data reduction for multichannel music
It seems likely therefore that multichannel music formats on DVD will optionally allow the use of data reduction techniques in order to add more flexibility into the format. Both lossless and lossy (perceptual) data reduction techniques have been proposed, the former effectively increasing the playtime but not the maximum number of channels, the latter permitting an increase in playtime and number of channels.

Lossless compression, such as entropy coding, is capable of reducing the average bit-rate of digital audio data by a factor of two. In order to increase this ratio to around three more sophisticated algorithms are required operating in the frequency domain or within sub-bands. The DVD hardware is fundamentally capable of delivering variable rate data to an audio decoder, constrained by a maximum audio data rate of approximately 6.5 Mbps. However since the peak data rate of a lossless audio coder may momentarily exceed the equivalent linear PCM data rate (due to management overheads in the coded bit stream), the maximum number of losslessly coded audio channels is equal to (or less than) the maximum number of linear PCM channels. For losslessly coded audio with a sample rate of 96 kHz or at a higher precision than 16-bit the average compression ratio may rise but the peak data rate will increase proportionately to the linear PCM bit-rate.

Any variable rate audio coding scheme (lossless or lossy) is also incompatible with the variable rate MPEG-2 video coding technique currently used on video DVD i.e. the audio and video could not co-exist on the same disc. Lossless compression does not allow the sonic performance of DVD to be extended beyond that which can be reached using linear PCM coding, but will increase the playing time of a DVD disc.

On the other hand the use of a perceptually transparent audio compression algorithm, operating at a fixed bit rate of around 256 kbps per channel at 48 kHz (512 kbps at 96 kHz), can increase the performance of multichannel music on DVD beyond the constraints imposed by multichannel linear PCM coding, and also deliver a greater number of audio channels. An appropriate perceptual coder would allow any multichannel format up to maximum of eight channels to evolve easily within either the audio or video DVD platform. This is illustrated in table 2. If the linear PCM stereo track were to operate at a sampling rate of 96 kHz the multichannel formats would be unchanged at 48 kHz, and be reduced to a maximum of six channels at 96 kHz. There is also sufficient data bandwidth in many of the modes for optional MPEG-1 video data.

Table 2 Multichannel audio modes assuming mandatory 2-ch PCM @ 20-bit @ 48 kHz (1920 kbps)

sample rate [kHz]	word length [bits]	maximum number of 'perceptually coded' audio channels for multichannel format [256/512 kbps per channel]								multichannel bit rate [kbps]	total bit rate [kbps]
48	16	x	x	x	x	x	x	x	x	2048	3968
48	20	x	x	x	x	x	x	x	x	2048	3968
48	24	x	x	x	x	x	x	x	x	2048	3968
96	16	x	x	x	x	x	x	x	x	4096	6016
96	20	x	x	x	x	x	x	x	x	4096	6016
96	24	x	x	x	x	x	x	x	x	4096	6016

Multichannel CD player

At these reduced bit-rates the CD platform also has great potential as a high quality multichannel audio player for the consumer, albeit without any stereo linear PCM audio track and with a maximum playing time of 74 minutes. The maximum bit-rate is 1.411 Mbps which allows up to six full bandwidth data reduced audio channels at 20-bit resolution and sampling rates of either 44.1kHz or 48 kHz. The use of CD media in these multichannel modes has been successfully demonstrated by Tom Jung of DMP Records using the DTS Coherent Acoustics algorithm. More details of these experimental recordings are given below, and some technical specifications of the DTS algorithm are presented in appendix 1.
It appears therefore that both the CD and DVD platforms are capable of delivering multichannel music to the consumer provided appropriate digital audio data reduction techniques are employed.

Experimental multichannel music recording for 5-channel and 6-channel playback

For the past couple of years Tom Jung of DMP Records has been recording and mixing music with the intention of reproducing it in a multichannel as opposed to a stereo format. This work has included both 4:2:4 and 5:2:5 matrix encoders, and more recently the discreet DTS Coherent Acoustics digital audio encoding technology.

Within the CD platform the DTS algorithm allows the reproduction of up to six full bandwidth channels of audio at a sampling rate of either 44.1 kHz or 48 kHz and at 20-bit precision. The six channels can be configured in a number of ways. These include a 4+2 format with essentially two front stereo pairs of speakers and stereo surrounds, a 3+3 circular format which tends to stabilize the stereo surround signal, and more recently a 3+2+1 format with the sixth channel vertically above the listener pointing down to convey height information. This last format has an advantage in that the more standard 3+2 format can be extracted and played without any mixdown. Owing to the importance of the home theater market the five channel (3+2) format has been predominantly used for reproduction, with the omission of the 0.1 low frequency effects channel. The usefulness of a separate bass channel for music recordings is not apparent.

In general the recordings were made with matched microphones feeding into multiple 20-bit A-D converters (up to 16-channels) locked to a common low-jitter clock source. The audio data was routed through a Yamaha digital mixing console to multiple 8-track digital audio tape recorders. The number a type of microphones used and their placement depended on the recording. Individual sections of the big band were recorded in stereo i.e. stereo trumpets, stereo trombones. These stereo pairs were assigned i mixdown to left and right walls (trombones left-front left-rear, trumpets right-front and right-rear). The saxophones were recorded on three channels and assigned accross the three front channels. The rhythm section was miked separately and assigned accross the front three channels. Two omnidirectional ambient microphones were used back in the room and assigned to left and right surround. An additiona directional microphone was used to record ceiling reflections and assigned to the sixth overhead channel The combination of the two rear surrounds and the overhead creates an interesting 3 dimensional ambie soundfield.

The intent during the jazz band mix was to wrap the band around the listener like a horse-shoe using stereo imaging down the side-walls. In general only ambient and strong reflective signals were reproduc from the surround channels apart from some deliberate and momentary placements of individual instruments in the rear. The center channel was used to create a strong frontal presence with further imaging between the center and front channels.

Monitoring was through matched mid-sized speakers with matched amplfiers, arranged in the standard equidistant 5-channel configuration, the speakers pointed directly toward the recording/mixing position. The rear speakers subtended an angle of approximately 90 degrees. When in use the sixth channel was positioned about 8 feet above and slightly in front of the mixing position, pointing straight down. Matchi the output levels of all the speakers has proven problematic due to inconsistencies in the output levels of pairs of stereo D-A converters and stereo amplifiers. Multi-channel digital converters with individual channel gain controls would be an advantage, and would also tend to lower the clock jitter problems associated with multiple stereo A-D and D-A units.

Encoding and decoding

The DTS 6-channel encoder operates in real time and allows the six channels to be monitored through the encoding / decoding cycle. The input to the encoder is via three AES-EBU digital audio channels which are normally phase aligned. The compressed data output is on a single AES-EBU channel, clocked synchronously (but delayed) by the digital audio inputs, and can be recorded on any digital audio recorder such as DAT, CD-R or a digital audio workstation. The DTS decoder receives compressed audio data via optical Toslink or single ended RCA jack, and outputs the six channels of de-compressed audio in both analog and digital forms.

The encoder/decoder is therefore fully digital and capable of operating at 24-bit precision. The simplest portable recording format for the 5-channel or 6-channel 20-bit source material is on a Tascam DA-88 wit a Prism Sound 20/24 digital interface unit which connects directly to the encoder. The ability to edit compressed audio is very limited due to the framing structure used in the bit stream. Edits must occur at frame boundaries and whole frames must be discarded or inserted.

Demonstrations of these five and six channel recordings, played back from a regular CD player into a DTS Coherent Acoustics decoder, will be given during the workshop.

Conclusion

Recording music for multi-channel presentation is challenging. Increasing the number of speakers and standardizing their placement in the listening room is the first stage and, compared to stereo, will dramatically increase the number of variables and decisions for producers and recording engineers.

The movie industry has provided a useful starting point for discussion with the 5.1 format, but the applicability of this for multichannel music has not been demonstrated. Given the scarcity of multichannel music recordings, the shortage of experimental data concerning discreet multichannel music formats, and the extra degrees of artistic freedom that the additional channels provide, it would be pre-mature to limit a multichannel music standard to those linear PCM formats that fit within the current DVD specification. In addition higher sampling rate audio (perhaps 64 kHz) at longer word lengths (at least 20-bit) are within reach of the consumer. A multichannel music standard that does not allow audio performance to move in these directions may not endure.

Appendix 1

Technical Overview of DTS Coherent Acoustics

Design objectives

The DTS Coherent Acoustics compression algorithm was designed from the outset to perform at studio quality levels i.e. 'better than CD', and in a multichannel format was intended to facilitate a major advance in the quality of audio reproduction in the home in terms of fidelity and sound stage imagery.

Another primary objective was that the compression algorithm should be broadly applicable and therefore flexible. Multimedia applications have restricted data bandwidths and therefore demand a 5.1 channel mode operating at 384 kbps or less. Professional music applications involve higher sampling rates, longer word lengths, multiple discreet audio channels, and increasingly demand lossless compression. All of these features have been accommodated in Coherent Acoustics.

The final important objective was to ensure that the universal decoder algorithm was relatively simple, and future-proofed. This would ensure cost effective consumer decoding hardware today, and yet allow consumers to benefit from any future improvements realized at the encoding stage.

Principal Encoding Processes

Coherent Acoustics is essentially a perceptually optimized differential subband coder. Figure 1 illustrates the main functional blocks of a single channel encoder. By combining differential coding and psycho-acoustically modeled noise-mask thresholds, the coding efficiency at very low bit rates can be enhanced thereby lowering the bit-rate at which subjective transparency is achieved. Multiple channels of audio are coded by allocating bits over all channels, either at fixed rates or adaptively.

A multirate filter-bank is used to split each single channel PCM source signal into 32 bands of equal bandwidth. This choice of filter combines the advantages of a high theoretical coding gain and excellent stop-band attenuation, with a low computational complexity.

Differential coding occurs within each subband, which removes most of the objective redundancy from the audio signal. In parallel psychoacoustic and transient analyses are performed on the uncoded signal to determine perceptually relevant information. The results are used to modify the main differential coding loop operating on each subband signal. In multichannel formats the bit allocation routine operates over all the coded channels, and adapts over time, frequency and channels to optimize audio quality.

Principle decoding processes

The decoding algorithm is quite simple compared to the encoder (figure 1) and does not involve calculations that are of fundamental importance to the quality of the decoded audio, such as the bit allocation. This ensures that future improvements can be made to Coherent Acoustics by modifying the encoding algorithm only, and that these improvements will then be realized by all decoders without any software or hardware change.

After synchronization the decoder unpacks the compressed audio bit stream, detects and if necessary corrects transmission induced errors, and demultiplexes the data into individual audio channels. The subband differential signals of each audio channel are re-quantized to PCM signals, and each audio channel is then inverse filtered to convert the signal back to a time domain signal. The DSP functional block allows operations to be performed on either the subband signals or on the time domain signals, on individual channels or globally across all channels. These functions include for example down-mixing, dynamic range control, re-equalization and differential time delays.

Coherent Acoustics Features

ñ 1 to 8 channels of multiplexed audio
ñ Sampling rates from 8 kHz to 192 kHz per channel
ñ 16-bit to 24-bit audio word length (138 dB)
ñ Compression ratios from 1:1 to 40:1
ñ Total data rate operating range from 32 kbits/sec to 4096 kbits/sec
ñ Lossless coding mode (variable data rate)
ñ Linear PCM decoding mode
ñ Down mixing from n coded channels to n-1, n-2, n-3 ...etc. output channels
ñ Down mixing from 5.1 discreet to matrixed stereo L_T R_T
ñ Embedded dynamic range control
ñ Re-equalization of all channels independently
ñ Sample accurate synchronization of audio to external video signals
ñ Embedded time stamp and user data
ñ Future proofed decoder

Multiple multiplexed audio channels

The specification allows from one to eight channels of compressed audio to be demultiplexed and decoded from a single data stream. The bit allocation for each multiplexed audio channel from the total data bandwidth, may be fixed or vary dynamically, depending on the demands of the application and the complexity of the encoder.

The total multiplexed compressed audio data rate can vary from 32 kbps up to 4.096 Mbps, depending on a number of application-defined parameters. These constraints would include, for example, the number of audio channels coded, the complexity and latency of the encoding process, the sampling rate of the source PCM digital audio, and the data buffer size at the decoder. This wide data bandwidth operating range allows Coherent Acoustics to find widespread use in diverse audio and audio/video applications. These range from telephone-grade voice audio at low data rates, 5.1 channel motion picture sound tracks around 384 kbps, and multi-channel music formats operating at very high sampling rates and with extended audio precision at the highest data rates.

Sampling rate and audio word length

Allowed sampling rates vary from 8 kHz per channel up to 192 kHz per channel, giving audio bandwidths from 3.5 kHz to 90 kHz. For applications which use fixed sampling rate D-A converters interpolation and decimation filters are included in the algorithm to up- or down-sample the PCM audio to the standard 32, 44.1 or 48 kHz sampling rates. The encoding algorithm operates with 40-bit precision on 24-bit audio words. Using computer generated 24 bit test signals, and with a 32-bit floating point decoder, the Coherent Acoustics algorithm can realize a dynamic range of up to 138 dB.

The higher sampling rates and extended word lengths of the algorithm are intended for next generation professional and high-end audio applications.[1] There is increasing use of 20-bit A-D converters in digital recording instruments, and a desire to increase the sampling rate to 64, 88.2 or 96 kHz for professional recordings. Coherent Acoustics provides a means of reproducing this higher quality audio in the home on current digital media such as compact discs or laser discs.

In addition to operating with lossy compression at fixed data rates from 32 kbps to 4096 Mbps, Coherent Acoustics is also capable of lossless compression in a variable rate mode, by forcing the coding error to be less than a fixed absolute value (i.e. less than +/- 0.5 LSB).

Encoding System Architecture

Input Channels

Up to eight discrete full-bandwidth digital audio channels can be input and coded as a single multiplexed data stream. An additional low frequency effects (LFE) can also be input to the coder in any mode. The pre-defined channel configurations are shown in table 1. Custom channel configurations are also possible as well as future strategies which can be uploaded from the player hardware.

Table 1
Audio coding modes

1-ch	A	(mono)
2-ch	A + B	(dual mono)
2-ch	L + R	(stereo)
2-ch	(L+R) + (L-R)	(sum-difference)
2-ch	Lt Rt	(total)
3-ch	L + R + C	
3-ch	L + R + S	
4-ch	L + R + C + S	
4-ch	L + R + SL + SR	
5-ch	L + R + C + SL + SR	
6-ch	L + R + CL + CR + SL + SR	
6-ch	Lf + Rf + Cf + Lr + Rr + Cr	
6-ch	L + R + C + SL + SR + OH	
7-ch	L + CL + C + CR + R + SL + SR	
8-ch	L + CL + C + CR + R + SL1 + SL2 + SR1 + SR2	
8-ch	L + CL + C + CR + R + SL + S + SR	

Sampling Rates

Permissible sampling rates are based on multiples of 32, 44.1 and 48kHz. This allows for straight forward decimation and interpolation to be carried out in the decoder in the event that the playback hardware sampling rate does not match the bit stream rate. Table lists the optional sample rates.

Table 2
Source PCM sampling rates

8kHz
16kHz
32kHz
64kHz
128kHz
11.025kHz
22.05kHz
44.1kHz
88.2kHz
176.4kHz
12kHz
24kHz
48kHz
96kHz
192kHz

ce PCM Word Length
encoding process allows for the coding of PCM signals with a dynamic range in excess of 144dB and
e facilitates PCM word lengths up to 24-bits with out truncation. Further more, the source word length
gged to the decoder in order to permit the implementation of noise shaping if truncation is
ssitated during playback.

: 3
ce PCM word length

ts
ts
ts
ts
ts
ts
ts

ding Bit Rates
: 4 lists the possible bit rates for the combined multi-channel encoded bit stream. Generally the lower
tes are intended for low sampling rate applications while the high data rates, specifically 3.072 and
iMbps are intended for sampling rates above 48kHz.

: 4
ding bit rates

)s
)s
)s
)s
ps
ps
ps
ps
ps
ps
ps
ps
ps
ps
ps
ps
ps
bps
bps
bps
bps
bps
kbps
bps
bps
bps
bps
bps

3840kbps
4096kbps
Variable
Lossless

Coherent Acoustics Coding Strategy

Detailed block diagrams of the subband ADPCM process at the encoder and decoder are illustrated in figures 2 and 3 respectively.

Input PCM Analysis Framing
The PCM analysis frame defines the number of contiguous input samples over which the encoding process generates an output frame. Five alternate encoding PCM frame sizes are permissible depending on the sampling frequency and the combined bit rate of the application. These are; 256, 512, 1024, 2048 and 4096 samples for each channel. Generally the larger window sizes are reserved for low bit-rate applications where the associated improvement in coding gain counteracts the falling number of bits available to code the audio . The maximum PCM window verses sampling frequency and bit rate is given in table 5. For example, encoding 48kHz sampled PCM digital audio channels at a combined rate of 384kbps allows the encoder to operated at any of the five frame sizes, whereas at a rate of 1.536Mbps the choice is reduced to either 256, 512 and 1024 samples per audio channel. The frame size limits in table 5 are imposed in order to maintain a maximum ceiling on the decoder input buffer size. For example, the maximum decoder input buffer size for sampling rates of 48, 96 or 192kHz can not exceed 5.3kbytes, irrespective of the number of audio channels present in the bit stream.

Table 5
Maximum allowed audio window size vs sampling frequency and bit-rate

	8 / 11 / 12 [kHz]	16 / 22 / 24 [kHz]	32 / 44.1 / 48 [kHz]	64 / 88.2 / 96 [kHz]	128 / 176 / 192 [kHz]
0 - 512 kbps	1024	2048	4096	-	-
512 - 1024 kbps	-	1024	2048	-	-
1024 - 2048 kbps	-	-	1024	2048	-
2048 - 4096 kbps	-	-	-	1024	2048

32-band Filter-bank (@fs=8-48kHz)
For sampling rates for 48kHz and below each PCM audio channel is split directly into 32 uniform subbands prior to encoding. A choice of two polyphase filterbanks with differing reconstruction properties are provided which allow for a trade off between subband coding gain and reconstruction precision. The choice of filterbanks is indicated to the decoder via a flag embedded within the encoded bit stream. In the case of low bit-rate coding the coding efficiency is enhanced by deploying high subband gain filter-banks i.e., those which exhibit high stop band rejection ratios. These filters are, by definition, prone to non prefect (amplitude) reconstruction (NPR) distortion at peak input levels such that the decoded PCM cannot match the input data bit-for-bit even at high bit rates. Hence a "perfect reconstruction" (PR) filterbank in included which is suitable for high-bit rate applications or lossless coding applications. The critical specifications for both filterbanks are shown in table 6.

Table 6
Filter bank specifications

Type	Taps	Transitional BW [Hz]	Stopband rejection [dB]	Ultimate rejection [dB]	Reconstruction resolution [dB]
NPR	512	300	110	120	90
PR	512	350	85	90	145

LPC analysis and vector quantization of predictor coefficients
A key component of the Coherent Acoustics coding process involves adaptive predictive coding, or ADPCM, which may optionally operate independently in all 32 subbands of each audio channel. Fourth order forward adaptive linear prediction is deployed where the optimal predictor coefficients are calculated over a window of either 8, (frame=256) 16 (frame=512) or 32 (frame= 1024 or greater) subband PCM samples. Each set of four predictor coefficients are quantized using a 4-element tree-search 12-bit vector code book prior to transmission, were each VQ element consists of a 16-bit integer (VQ table=32kbytes). For example, in a 4096 PCM sample coding frame where the signal is decimated to 32 subbands of 128 PCM samples, the subband predictor coefficients are updated and transmitted to the decoder four times for each incoming frame.

Subband Difference Signal Estimation
Once the optimal prediction coefficients have been calculated and quantized for each subband, a first pass difference signal is generated by running "estimation ADPCM" routines which assume zero quantization error in each subband. Using the estimated difference signal an initial analysis of both the prediction gain and transient behavior of the subband difference signal is made.

By comparing the variance of the difference signal to that of the subband signal the prediction gain can be estimated over each analysis window. If the prediction gain is not sufficiently positive (taking into account the prediction coefficient transmission overheads and prediction gain loss due to quantization error feedback) then the prediction process is disabled in that subband (coefficients set to zero) for the period of that analysis window (nominally 32 subband samples). The use, or otherwise, of the predictors in any subband is indicated directly to the decoder via "predictor mode" flags (PMODE) embedded in the data stream. Prediction coefficients are not transmitted for any subband analysis window for which the predictor mode flag is off.

In this way the prediction process is dynamically activated in any subband if the accuracy of the prediction sequence is deemed to provide a realistic reduction in quantization noise at a given bit rate over the period of that subbands analysis window.

Differential Transient Analysis
For those subbands which do not exhibit a prediction gain, the estimated difference signal samples for those periods are overwritten by the original subband PCM samples for those bands. The updated estimated difference signal is then re-analyzed for the presence of transients which could lead to audible pre-echo artifacts.

The "transient modes" are calculated separately for each subband estimated difference signal and indicate the number of scale factors required to be calculated over the transient analysis window. This window is similar to that of the prediction analysis i.e., 8 samples (frame=256), 16 samples (frame=512) or 32 samples long (frame=1024 or greater). A transient is defined as a signal which transitions rapidly between a low amplitude phase and a high amplitude phase. Due to the fact that scale factors are averaged over a block of difference samples, if a transient is present, the calculated scale factor can be excessive in relation to the low level samples which preceded it, possibly leading to pre-echo at low bit-rates.

To alleviate this possibility, the position of the transient is located within the analysis window and one four transient mode values (TMODE) assigned to each subband window. The process is illustrated in figure 4, for the case of where the sub-buffer comprise 8 differential samples. For example, for a cod frame of 4096 samples per channel, the transient mode analysis is conducted over 4 analysis buffers (windows) of 128 differential subband samples. Each 128 sample window is split into 4 sub-blocks of samples (5.3ms @48kHz) and the RMS scale factors calculated for the various combinations or group sub-buffers (TMODE=0,1,2 and 3). For transients which occur in the first sub-buffer the mode is alw

Two bit TMODE flags for each analysis window in each subband are embedded in the bit stream and used directly by the decoder to properly unpack the embedded scale factors.

Scale Factor Generation
Scale factors are calculated for each grouping of 8-sample sub-buffers as indicated by TMODE for eac subband estimated difference signal window. When TMODE is zero a single scale factor is calculated the entire window. The calculation can be based on either sample Peak Amplitude or RMS depending c the application. As a result of the transient analysis method the effective scale factor time resolution ca be as low as 5.3ms while only requiring the transmission of a maximum of two scale factors over a 21 window (@48kHz).

Depending on the application, scales factors are quantized logarithmically using either a 64-level (2.1d step) table or a 128-level (1.3dB step) table, which allow for the dynamic tracking of audio over range 134dB (22-bits) and 166dB (27-bits) respectively. Scales factors for each subband of each audio channe are transmitted directly to the decoder and converted back to the linear domain using a simple look-up table at the decoder. The choice of quantization tables is also embedded in the bit stream for each analysis frame.

Use of dynamic transient analysis in combination with the scale factor generation process within each subband, pre-echo artifacts are conveniently mitigated at low bit rates with minimal increases in side information overheads. Moreover, since the transient data is conveyed directly to the decoder, the transient analysis can be continually improved with time without impacting the decoder.

Psychoacoustic Analysis
Traditionally psychoacoustic analysis for low bit-rate subband coding has generated short-term signal-t mask ratios (SMR) which determine the number of bits required for adaptive PCM encoding within ea subbands (MPEG Layer I and II). However, for predictive subband coding the SMRs must be modified reflect the degree of prediction gain obtained in each subband ADPCM process since it is the differenc signal which is to be quantized as opposed to the actual subband sample. Certain care must be excercised however, since large prediction gains can lead to the SMRs becoming negative and possibly resulting in zero bit allocations. Generally speaking, subbands whose SMRs are initially negative can receive a zero bit allocation, whereas subbands whose SMRs become negative after modification must retain at least a minimum bit allocation.

Adaptive Bit Allocation
Once the scale factors have been generated, the quantization noise level is set by selecting the number c levels to be used by the differential quantifier (the samples are first normalized by the scale factor). At bit-rates acceptable levels of quantization noise in each subband may be determined by either the SMR values generated either by the psychoacoustic analysis directly or from SMR values modified using the subband prediction gains, or a combination of both. Alternatively, at higher bit rates, a combination of SMR and differential minimum mean squared error allocation is possible. Furthermore, at lossless or variable rate coding mode the bit allocation can continue to accumulate until the quantization noise lies below some pre-determined threshold i.e. half a LSB of the source PCM audio, in the case of lossless coding.

oice of up to 28 mid-tread quantifiers (minimum 0-levels, maximum 16,777,216-levels), are available de the differential subband signals within the ADPCM process. Depending on the bit-rate of the cation the bit allocation indexes are transmitted directly to the decoder as either 4-bit or 5-bit words. of the four bit word reduces side information overheads but limits the choice of quantifiers to the first en. This would be appropriate for low bit-rate applications. The bit allocation index word length is mitted to the decoder via a flag each frame.

gle bit allocation index is transmitted for each subband differential signal analysis window, (up to two factors can be transmitted for the same period if TMODE is non-zero). Since the bit allocation es are sent directly to the decoder, the bit allocation process can be continuously improved over time out impacting the decoder.

Frequency Coding

ry low-bit rate applications it is possible to improve the overall reconstruction fidelity by coding only a nation of high frequency subband signals from 2 or more audio channels, as opposed to coding the frequency subbands separately. Joint frequency coding indexes are transmitted directly to the der and indicate which subbands contain summed signals and which channels have been involved in summation process. Frequency joining is permissible in all subbands except in the first two. In low te high-quality applications, typically frequency joining would be limited to subbands in the region 10-z. In medium to high bit-rate applications the feature would be disabled altogether.

ard ADPCM

the prediction coefficients, scale factors and bit allocation indexes have been determined the nd samples are encoded using ADPCM. The resulting quantifier level codes can be sent directly to it stream multiplexer in medium to high bit-rate applications, or they can be post-processed by ing to variable length code books for further bit rate reduction.

able Length Coding

the statistical distribution of the differential quantifier codes is significantly non-uniform further vements of up to 20% in the coding efficiency may be realized by mapping the code words to a ble length "entropy" code books. Variable length coding is appropriate when operating at low-bit rates vo reasons. One, since the coding windows may be up to 4096 sample per channel at low bit rates, veraging effect of a large number of differentially code words flattens the variance of the frame bit Two, although the unpacking complexity of variable length codes is significantly greater than fixed codes, there will be fewer bits to unpack at low-bit rates. Hence the unpacking computation is vely constant between high and low bit rates. Depending on the size of the linear quantifier, a number tistically different entropy tables are available for mapping purposes. The codes from the table which ces the lowest bit rate are used to replace the fixed differential codes, and these are sent to the plexer. Flags indicating which table has been selected are transmitted alongside the codes to ate proper decoding. If the bit rate of the variable length codes is not less than the original fixed codes, then the fixed codes are transmitted instead.

der Bit-rate iteration

case that variable length codes have been elected to re-map the differential codes, it is possible that combined frame bit usage across all audio channels exceeds the bit rate of the application. s case an iterative approach is deployed in the encoder whereby certain scale factors (normally high ency) are increased incrementally in order to force the entropy mapping process to progressively use er code word lengths. For each iteration the ADPCM process and entropy mapping is repeated and tal bit rate recalculated. In practice this process rarely continues for more than one or two iterations.

Side information encoding

Although the transient modes (TMODE), scale factors and bit allocation indexes can be transmitted directly, in low-bit rate applications (below 100kbps/ch) these combined side information overheads can become a significant portion of the total bit rate and will inevitability begin to limit the quality of the decoded audio below these rates. Typically an overhead of approximately 14kbps/ch can be expected for bit-rates in the region 64-100kbps/ch. By re-mapping the side information using variable length code books in a similar fashion to the differential subband codes, the average side information at these bit rates is reduced by approximately 3kbps/ch (21%) to about 11kbps/ch.

Low Frequency Effects Channel

The low-frequency audio channel is optionally available with all audio coding modes. The effects channel is derived by directly decimating a full bandwidth input PCM data stream using a choice of either 64X or 128X decimation digital filters. These filters exhibit bandwidths of 150Hz and 80Hz respectively. The decimated PCM samples are coded using an 8-bit forward adaptive quantifier. To reconstitute the PCM channel at the decoder the same filters are used to interpolate back up to the original PCM sample rate.

Embedded functions

In order that multi-channel audio formats be compatible with standard stereo or mono playback systems. the decoder includes functions for down-mixing n-channels to n-1, n-2, ..., etc. This also includes the ability to down mix a discreet 5.1 channel sound track to a matrixed 2-channel L_T R_T version, compatible with current matrixed surround sound decoders.

The decoder includes pre-set mix coefficients, but alternative coefficient values can be embedded by the encoder in the audio stream or input to the decoder through a serial interface. This allows program providers great flexibility in determining the optimum mix parameters for particular audio material. Channel mixing is possible either in the frequency domain or time domain, depending on the complexity of the decoding hardware.

Dynamic range control in all modes is facilitated through the inclusion of moving-average energy values for each compressed audio channel in the multiplexed compressed audio bit stream. Time stamp and user defined information for each channel may also be embedded in the bit stream.

By varying the block size of the compressed audio bit stream Coherent Acoustics also permits sample accurate synchronization of the audio to digital video signals at any sampling rate. This is of importance in multimedia and digital video playback applications where the compressed audio data blocks are not normally coincident in time with the beginning and end of the video frames. In this situation it is possible for the audio and video signals to loose synchronization with each other, especially if the video signal is being used for a timing reference. Time stamp information and user defined data may also be embedded in the bit stream.

High Sampling Rates

For source sampling rates greater than 48kHz the basic 32-band subband coding framework continues to encode the base-band audio signals (0-24kHz), however the higher frequencies are directed to additional 8-band (FsX2) and single-band (FsX4) side-chain ADPCM coders. The encoding process for 96kHz sampling is illustrated in figure 5. The audio spectrum, 0-48kHz, is initially split using a 256-tap 2-band decimation filter bank giving and audio bandwidth of 24kHz per band. The bottom band (0-24kHz) is encoded using the standard 32-band system. The top band (24-48kHz) however is split delayed and encoded in 8 uniform bands. Each of these subbands have a bandwidth of 3kHz and the ADPCM codes are packed within the oversampling array included in the main bit stream. By splitting the signal in this way, the decoder does not need to decode the audio codes associated with the 24-48kHz region to maintain compatibility with the bit stream. Figure 6 illustrates this process.

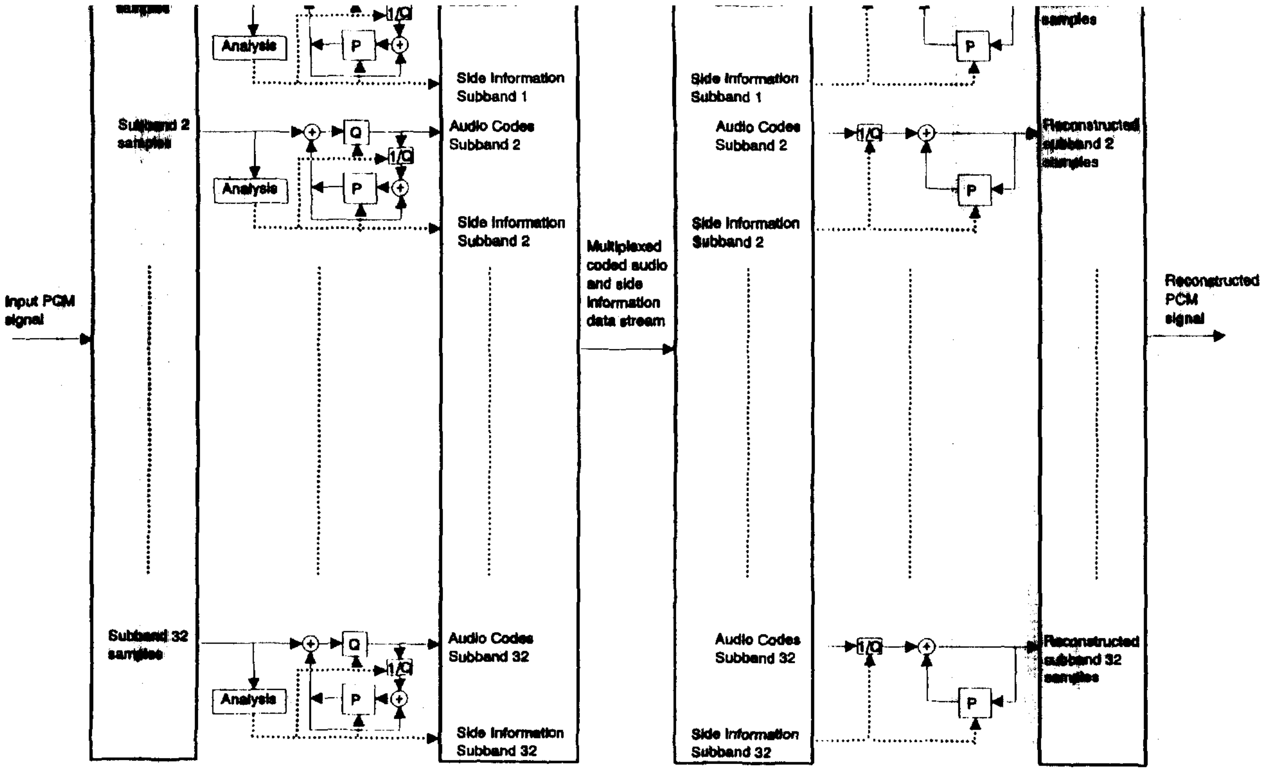


Figure 1 Basic principle of ADPCM encoding and decoding within 32-band Filter bank structure

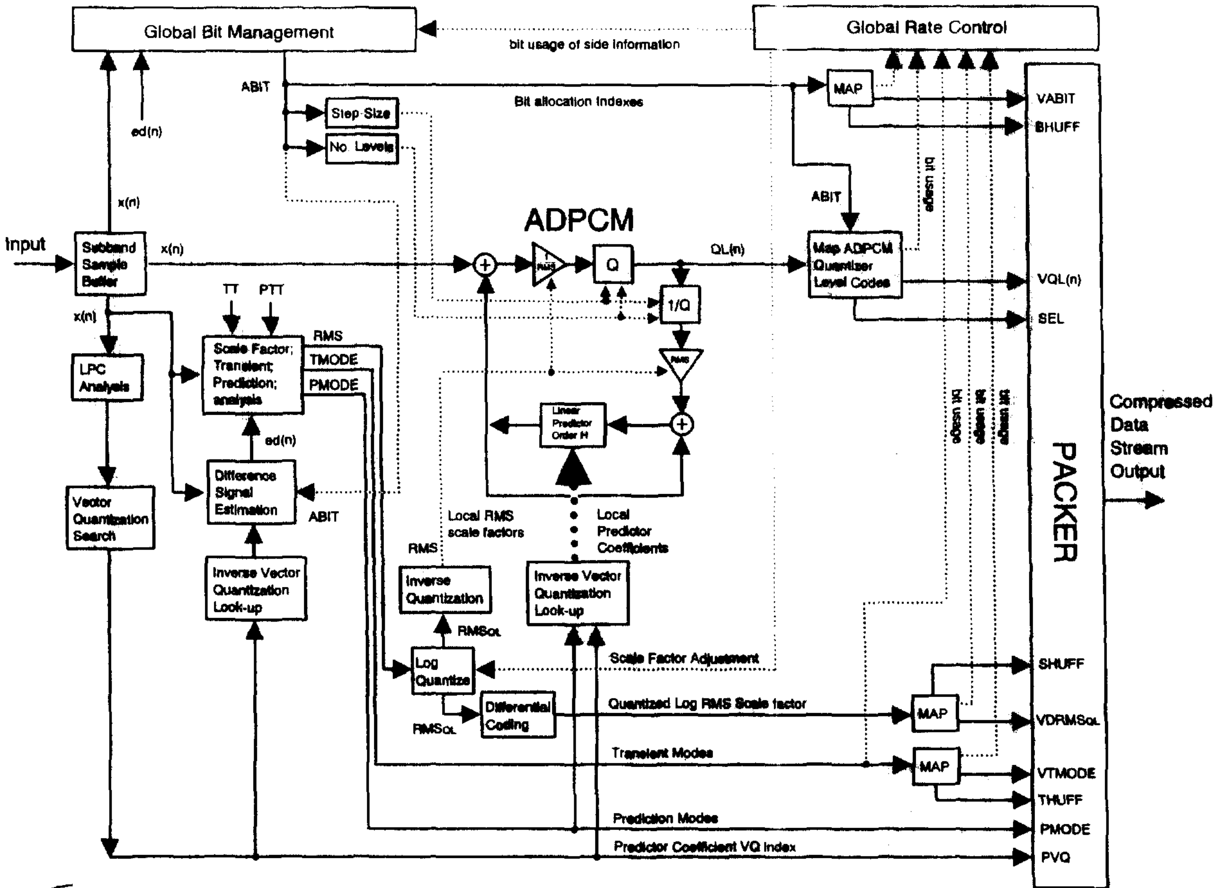


Figure 2 Main Processes involved in ADPCM Encoding of single Subband using Entropy encoding for differential subband samples and side information and RMS scale factors

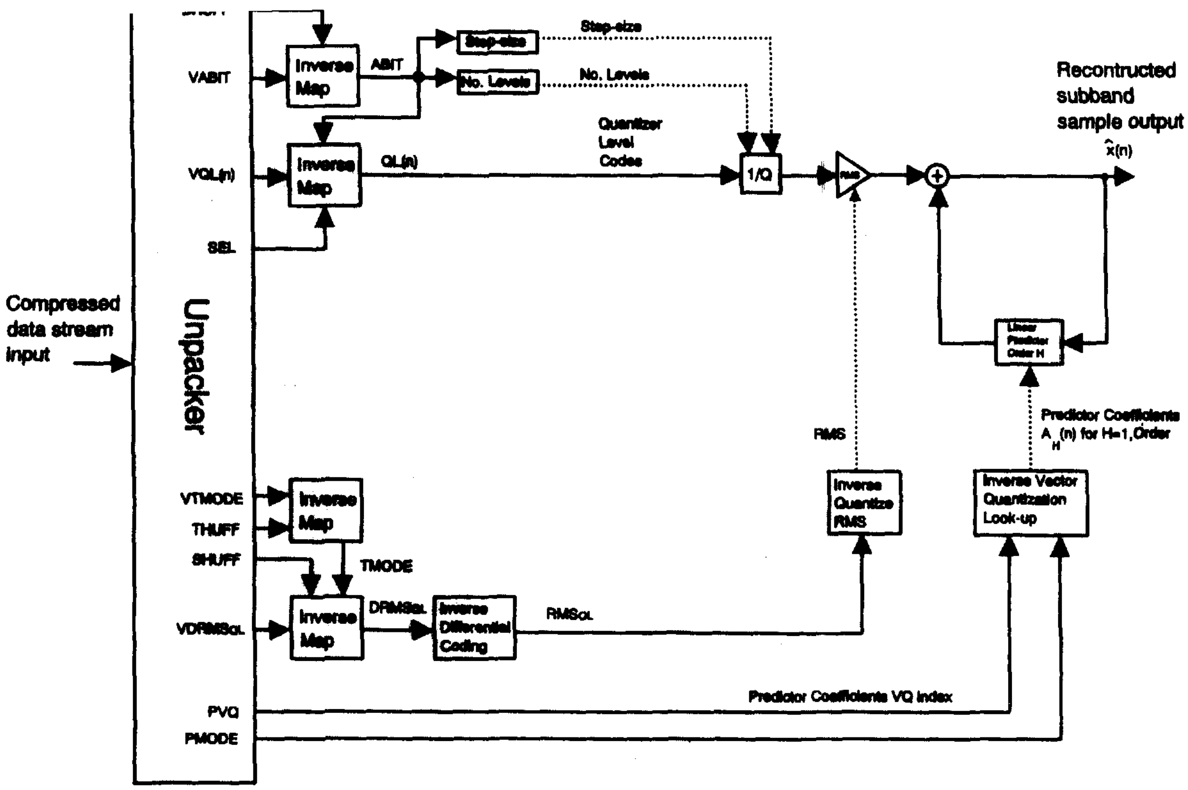
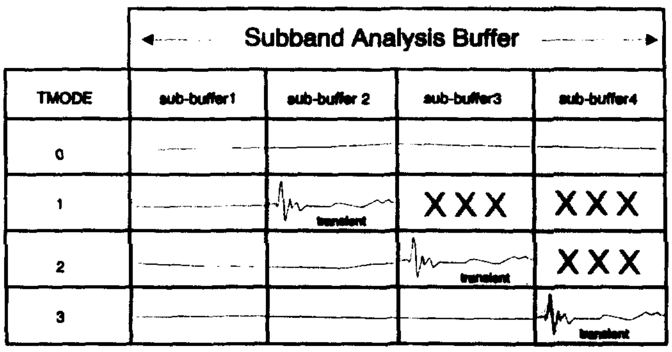


Figure 3 Main processes involved in ADPCM Decoding of single Subband using Entropy decoding for side information, differential codes and RMS scale factors



	Subband Analysis Buffer			
TMODE	sub-buffer1	sub-buffer 2	sub-buffer3	sub-buffer4
0	RMS1 or PEAK 1			
1	RMS1 or PEAK1	RMS2 or PEAK2		
2	RMS1 or PEAK1		RMS2 or PEAK2	
3	RMS1 or PEAK1			RMS2 or PEAK2

Figure 4 Calculation of Transient Mode (TMODE) for subband analysis buffer deploying 4 uniform sub-buffers and resulting grouping of sub-buffers for generation of scale factors values

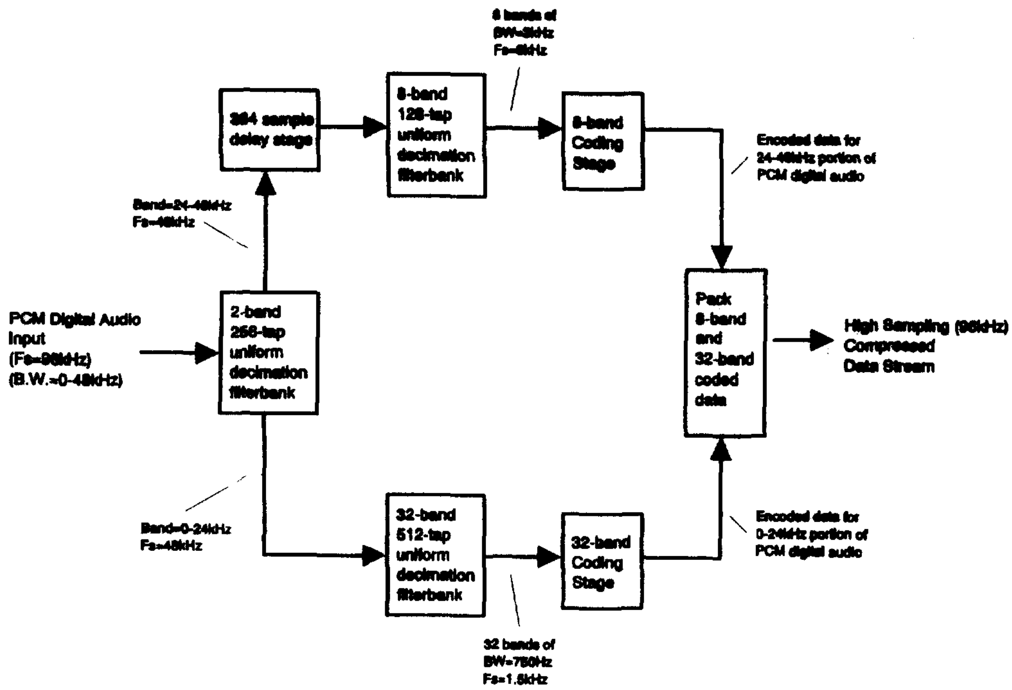


Figure 5 Encoding of High Sampled (96kHz) digital audio using 32-band coding for frequencies 0-24kHz and 8-band coding for frequencies 24-48kHz

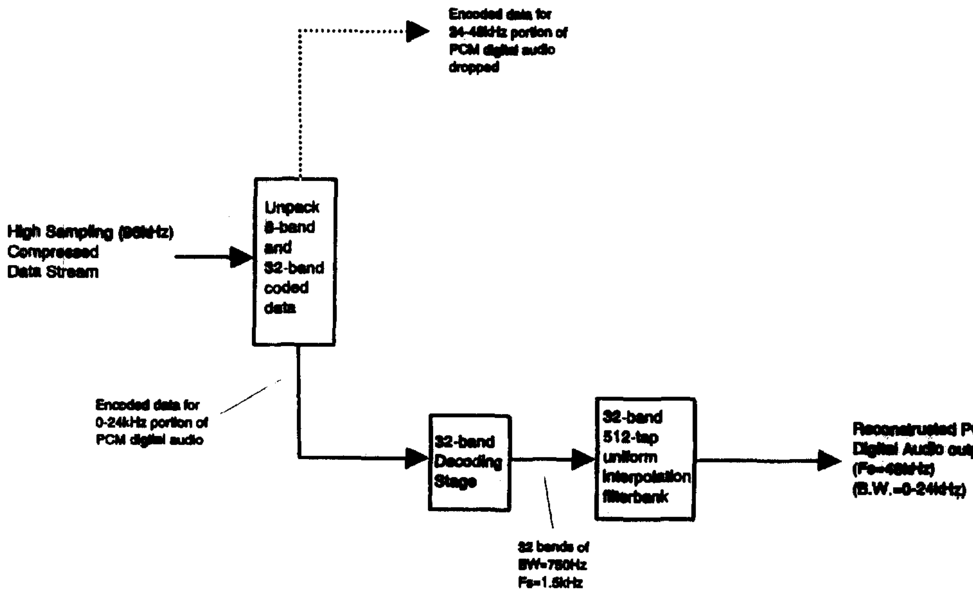


Figure 6 Decoding of High Sampled (96kHz) Compressed data stream using only 32-band decoding for 0-24kHz audio